

Fusing Spaceborne SAR Interferometry and Street View Images for 4D Urban Modeling

Yuanyuan Wang
Signal Processing in Earth Observation
(SiPEO)
Technical University of Munich (TUM)
Munich, Germany
wang@bv.tum.de

Jian Kang
Signal Processing in Earth Observation
(SiPEO)
Technical University of Munich (TUM)
Munich, Germany
jiang.kang@tum.de

Xiao Xiang Zhu
SiPEO, TUM, Munich, Germany
and Remote Sensing Technology Institute,
German Aerospace Center, Weßling,
Germany, xiao.zhu@dlr.de

Abstract— Obtaining city models in a large scale is usually achieved by means of remote sensing techniques, such as synthetic aperture radar (SAR) interferometry and optical image stereogrammetry. Despite the controlled quality of these products, such observation is restricted by the characteristics of their sensor platform, such as revisit time and spatial resolution. Over the last decade, the rapid development of online geographic information systems, such as Google map, has accumulated vast amount of online images. Despite their uncontrolled quality, these images constitute a set of redundant spatial-temporal observations of our dynamic 3D urban environment. These images contain useful information that can complement the remote sensing data, especially the SAR images.

This paper presents a one of the first studies of fusing online street view images and spaceborne SAR images, for the reconstruction of spatial-temporal (hence 4D) city models. We describe a general approach to geometrically combine the information of these two types of images that are nearly impossible to even coregister without a precise 3D city model due to their distinct imaging geometry. It is demonstrated that, one can obtain a new kind of city model that includes high resolution optical texture for better scene understanding and the dynamics of individual buildings up to the precision of millimeter retrieved from SAR interferometry.

Keywords—SAR, TomoSAR, structure from motion, optical images, 3D, 4D, urban model, fusion

I. INTRODUCTION

Reconstructing city models in a large scale is usually achieved by means of spaceborne or aerial remote sensing techniques, such as optical stereogrammetry and synthetic aperture radar (SAR) interferometry. For example, the latest global digital elevation model (DEM) was derived from the SAR sensors TerraSAR-X, and TanDEM-X; urban 3D building models in online maps are often derived from high-resolution spaceborne optical images such as WorldView and RapidEye images. They provide geodetically highly accurate information in large scale. However, these products are restricted by the characteristics of the sensor platforms, such as the revisit time and spatial resolution.

Within the last decade, the rapid development of online geographic information systems, such as Google map and Mapillary, has accumulated vast amount of freely available online images. Despite their uncontrolled quality, these images constitute a set of redundant spatial-temporal observations of

our dynamic 3D urban environment. They contain useful information that can complement the remote sensing images in 4D urban modelling. This is especially true for SAR images, because SAR images require an inevitable side-looking range-azimuth imaging geometry. This imaging geometry is intrinsically distinct from the perspective projection of optical images, which renders the SAR image difficult to interpret for human being. For example, Figure 1 shows a TerraSAR-X image and an online optical image of the German parliament building in Berlin, Germany. One can observe that neither the geometry nor the radiometry of the object in the two images is similar. Thus, it handicaps the application of SAR images in computer vision whose recent development greatly contributed to the 3D urban modelling, such as [1], [2].



Figure 1. Left: TerraSAR-X image of the Reichstag (parliament) building, Berlin, Germany, and right: an online street view image.

Apart from the authors' previous work, there has not been any study on fusing very high resolution spaceborne and ground level image of dense urban area. To this end, this paper demonstrates the possibility of fusing spaceborne SAR images and optical images from street level, such as social media images, for the reconstruction of a new type of spatial-temporal (hence 4D) building models. The new 4D building model will contain, in addition to the 3D geometry of the building, precise deformation measurement of building, and high resolution semantic information, derived from SAR and street view images, respectively. The challenge of this fusion task lies on the co-registration of the two types of images, which is nearly impossible to be achieved without a precise 3D model of the object.

In this paper, this challenge is tackled by the “SARptical” [3] framework we previously developed. It is a suite of sophisticated algorithms that performs coregistration and joint analysis of SAR and optical images. In the rest of the paper,

Section II will introduce the related work of 3D reconstruction using SAR and optical images, respectively, as well as the SARptical framework. Section III explains our experiment using images crawled from Google map and TerraSAR-X images. Section IV and V discuss the result and open issues.

II. BACKGROUND

A. Multipass SAR Interferometry

One of the most unique advantage of SAR is its capability for assessing long-term millimeter-level deformation as well as the 3D geometry of the scene over large areas via multipass SAR interferometry (InSAR). Multipass InSAR analyses the time series of scatterers' interferometric phases, to retrieve the corresponding geophysical parameters such as 3D positions and deformation parameters. Among the various multipass InSAR techniques, SAR tomography (TomoSAR) and its differential form (D-TomoSAR) [4]–[10] are the most popular ones for the abovementioned tasks in urban areas, because of their distinct abilities to separate multiple scatterers mixed within a pixel (so called “layover” in radar jargon), giving a truth 3D reconstruction of scene.

Several advanced D-TomoSAR methods have been developed, e.g. superresolution achievement in separating closely spaced scatterers based on compressive sensing technique [5], and fusion of SAR imaging geodesy [11] and D-TomoSAR to obtain absolute *Geodetic TomoSAR* point clouds [12]. Although the abovementioned multipass InSAR techniques are able to deliver precise 3D reconstruction and deformation estimates, a sufficient number of SAR images (at least 20) should be satisfied in order to achieve a reliable estimation.

B. Optical stereo matching and structure from motion

3D reconstruction from optical images has been investigated for a long time in photogrammetry, which is usually carried out by stereo matching [13]. Over the last decade, one of the major breakthrough in 3D reconstruction from optical images is structure from motion (SfM) [1], [2], [14]–[16]. It is designed to exploit large quantity of unstructured image data through matching the corresponding points among the images. Hence, it allows the recovery of both the camera parameters and the scene structure.

Moreover, 3D reconstruction of large area using hundreds of thousand social media images [1], [14] is also possible. Despite it is suitable to work with large quantity of unordered image, the availability of the social media images is only guaranteed for certain geographical areas.

C. The “SARptical” framework

Coregistering and joint analysing high-resolution SAR and optical images of dense urban areas was void before the development of the SARptical framework due to the extreme difference of the imaging geometry of SAR and optical images. SARptical is a general framework to perform such task. Algorithmically, it integrates the abovementioned two techniques, i.e. multipass InSAR, and optical SfM, as well as the 3D model coregistration algorithms. Its flowchart is

illustrated in Figure 2. In order to coregister these two types of images, independent 3D point clouds are reconstructed from the two datasets using D-TomoSAR and SfM with dense matching, respectively. Therefore, D-TomoSAR and SfM build up the core of the SARptical framework. The algorithms require a stack of tens of SAR images, and tens to hundreds of optical images, respectively, for the 3-D reconstruction.

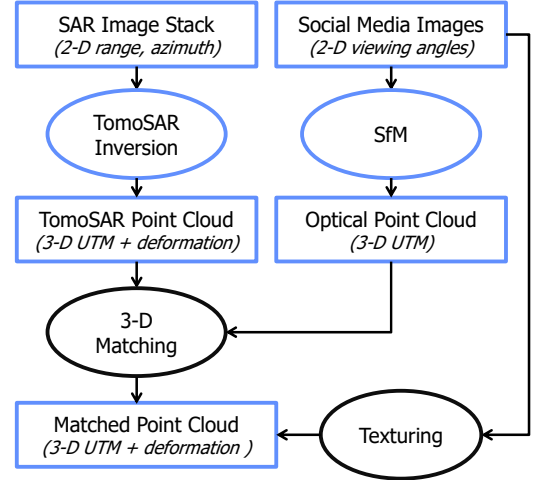


Figure 2. The flowchart of the SAR and social media image fusion algorithm. It matches the 3-D TomoSAR and optical point clouds, and then transfers the texture from optical images to the TomoSAR point cloud. This paper focuses on the 3-D matching and texturing which are shown in black ellipses. The coordinate system of each dataset in the flowchart is indicated by the italic text in the bracket. The TomoSAR point clouds contains the attribute of deformation parameters in addition to the 3-D position.

The essential steps in SARptical is summarized as follows.

- 3-D reconstructions: reconstruct 3-D point clouds from the SAR image stack and the optical images using D-TomoSAR and SfM, respectively.

The output of D-TomoSAR is a 4D point cloud (3D plus deformation parameters) of the scene. The absolute location of this point cloud is unknown due to a subtraction of a reference point whose position is unknown during the InSAR process. Such differential operation is often performed in multipass InSAR, in order to mitigate some common errors, such as orbit and atmospheric error.

SfM works on unordered optical images. The reconstructed 3D point cloud can have arbitrary scale and coordinate origin, because most of the social media images do not contain geo-tag. Ground control points are required to geo-localize the point cloud.

- 3-D matching: coregister the TomoSAR point cloud and the optical point cloud.

Both the SAR and optical point clouds are retrieved from oblique images from which rich façade points [17] can be maintained to a large extend. That is to say, these two point clouds share significant amount of common areas. Therefore, conventional point cloud coregistration methods, such as iterative closet point (ICP), can be exploited in this framework.

However, one must take into account the different modalities of the two point clouds. The accuracy of TomoSAR point cloud is very anisotropic. It has an extremely good

accuracy up to centimeter in the radar native range-azimuth coordinate [11], [18], but much poorer in its third coordinate *elevation*. A typical elevation accuracy for a stack of 50 TerraSAR-X images is about of 1~10m, which is about 100~1000 times worse than the other two dimensions. Transforming to Universal Transverse Mercator (UTM) coordinate, the error is mostly translated to the *vertical* and *east* direction. In addition, the accuracy is also not consistent over the whole point cloud, as the signal-to-noise ratio (SNR) of each pixel are very dynamic.

On the contrary, the optical point cloud is accurate in the normal direction of the façade and poor on the other direction, because most of the optical images used in this paper were acquired on the street level pointing to the façade. The accuracy of the optical point cloud derived from social media images will depend on the number of images, as well as their quality. In general, the optical point cloud has higher 3D accuracy than the SAR point cloud, because of their close range observation.

III. EXPERIMENTS

Several building blocks near Brandenburg Gate in the city of Berlin were chosen to be the experimental area in this paper. In the experiments, we made use of two stacks (ascending and descending) of TerraSAR-X high-resolution spotlight images acquired between February 2008 and March 2013. Each stack has about one hundred images. The optical images were crawled from Google map. We employed Google street view API as well as screenshot to captured the images from Google map.

A. Crawling optical images from Google map

Given the longitude and latitude of a point of interest, Google API can return the nearest street view image that looking towards the point of interest. Some other metadata, e.g. image size, and pitch degree, are also required. In the experiments, those two parameters were set as 512×512 pixels and 10 degrees, respectively. Some examples of the street view images of the buildings in the study area are illustrated in Figure 3. However, these street view images are often in low quality, such as image blurring due to privacy reasons and object occlusion due to trees and vehicles. Therefore, we also acquired some images by taking screenshots on Google map. About one hundred images around several blocks near Brandenburg gate in Berlin, Germany were crawled.

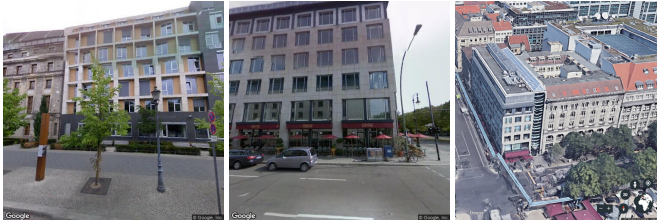


Figure 3. Examples of downloaded Google map street view images around the study region.

B. 3D reconstruction from SAR images

To perform D-TomoSAR inversion on the SAR image stacks, interferograms were firstly formed and coregistered

using the integrated wide area processor (IWAP) of German Aerospace Center (DLR) [19], [20]. The follow on D-TomoSAR inversion was performed by the tomographic inversion system *Tomo-GENESIS* [21] of DLR, which is based on the D-TomoSAR techniques [4], [5], [9].

In the D-TomoSAR reconstruction, the 3D position and the amplitude of seasonal motion of the scatterers in the scene were estimated. Shown in Figure 4, the top figure gives a view of 3D point cloud of Berlin and the bottom figure indicates its corresponding amplitudes of seasonal motion over the observation time. The variations of heights and deformations are described by different colour coding.

As mentioned in section II.C, the topography retrieved from spaceborne SAR images is moderately accurate (1 to 10m). Yet, rather precise deformation up to a millimeter-level can be achieved. It can be seen that most areas in Berlin generally undergo no deformation, except a few buildings, especially the Berlin central station (indicated by the red arrow in Figure 4) behaving significant periodic motion. This is mostly due to the natural thermal dilation of the metal structure caused by seasonal temperature variation. The black area in the figures is the region where no coherent signal was acquired over the time span of the images. These areas are usually vegetation and river.

C. 3D reconstruction from street view images

The area covered by the street view optical images is marked by the red rectangle in Figure 4. The area contains several building blocks. The result of SfM plus dense reconstruction of these building block is illustrated in Figure 5. The 3D point cloud and its dense reconstruction were reconstructed by VisualSFM [2], [22]. There are in total about 1.2 million points in the dense point cloud.

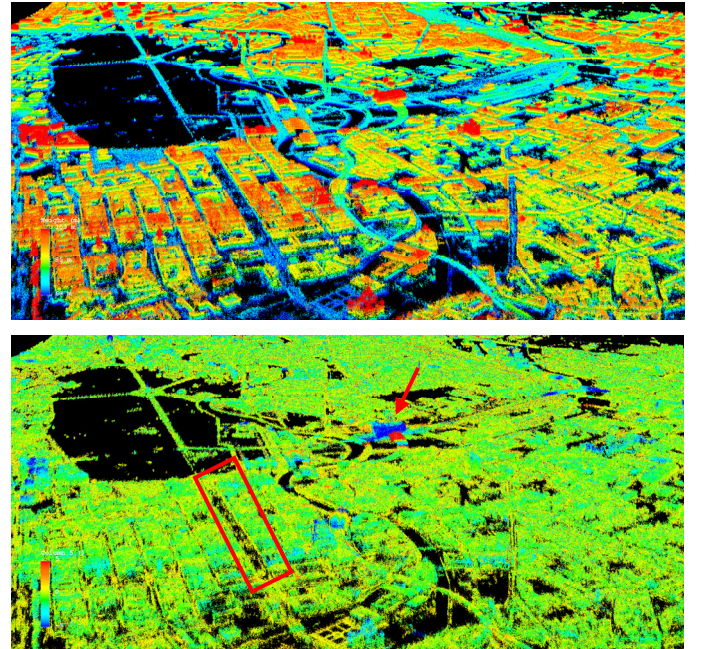


Figure 4. 3D point cloud (top) and the amplitude of seasonal motion (lower) reconstruction of Berlin. The unit are meter and millimeter, respectively. It can be seen that most buildings in Berlin is relatively stable,

since their amplitudes of seasonal motion are around zero. However, some buildings, such as the Berlin central station (indicated by the red arrow), have significant motion during the observation time. The area marked by the red rectangle is the study area covered by the street view optical image used in the experiment.

Compared to the TomoSAR point cloud shown in Figure 4, the 3D accuracy of the optical point cloud is better in one order of magnitude, i.e. around 10 to 50 cm. Moreover, the semantic information is more obvious in the optical point cloud than the TomoSAR point cloud. Building façade can be clearly distinguished. By fusing these two results, one can combine the unique advantage of both datasets, i.e. the precise deformation and the high-resolution semantic information.



Figure 5. Top: examples of the optical images captured from Google map, and lower: the reconstructed dense 3D optical point cloud of the study area. There are totally 1.2 million points in the point cloud.

D. Point cloud fusion

In order to perform the point cloud fusion, the two types of point clouds are coregistered using the iterative closest point (ICP) algorithm. Since the optical images do not have any geo-tag, it has an arbitrary origin and scaling. They are tackled by centering the point clouds and performing principle component analysis, respectively, before the ICP. A more precise scaling of the optical point cloud was estimated during the ICP. In the experiment, we cropped the TomoSAR point cloud to be roughly the same area as the optical point cloud, because of their great different in size. However, this is not necessary when coregistering an optical point cloud covering an area as large as the TomoSAR point cloud.

The coregistered optical and TomoSAR point clouds can be seen in the following figure where the color indicates different point clouds (green: TomoSAR, red: optical). As one can see that the optical point cloud is much denser than the TomoSAR one, because of its close range measurement.

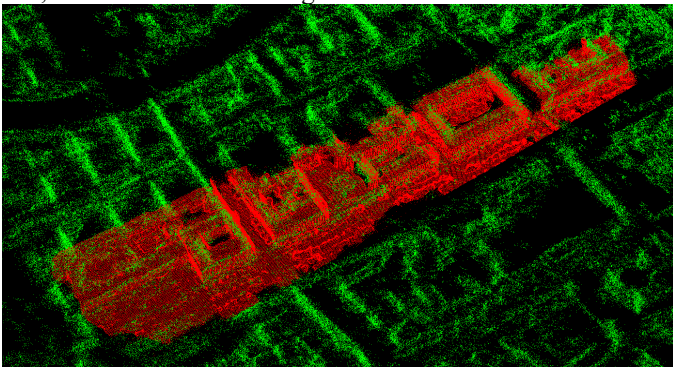


Figure 6. The coregistered optical (red) and TomoSAR (green) point cloud in UTM coordinate. The optical point cloud is much denser than the TomoSAR point cloud.

Once the two point clouds are coregistered, one can texture the TomoSAR point cloud with optical information, or vice versa. A surface was reconstructed from the optical point cloud using Poisson surface reconstruction [23]. Optical images were textured to the surface to generate a photo realistic 3D model of the scene. This can be seen in Figure 7. Based on this 3D model, the information from the TomoSAR point cloud, such as the intensity of the scatterer and the deformation, can also be textured to the optical 3D model. For example, Figure 8 shows the 3D building model textured by the scatterer intensity from the TomoSAR point cloud. Such information fusion is valuable for understanding the source of each bright scatterers appearing in SAR images. This helps the understanding of the complex interaction of the scatterers during the SAR image formation. In a similar way, the precise deformation obtained from TomoSAR can also be textured to the building model. This gives the urban planner an intuitive view of the deformation of individual building, even up to the detail of each floor.

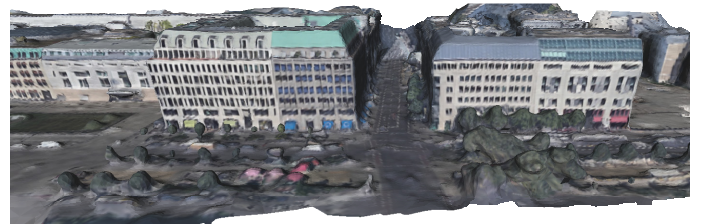


Figure 7. 3D surface reconstructed from the optical point cloud. Color from the street view images is textured to the surface.

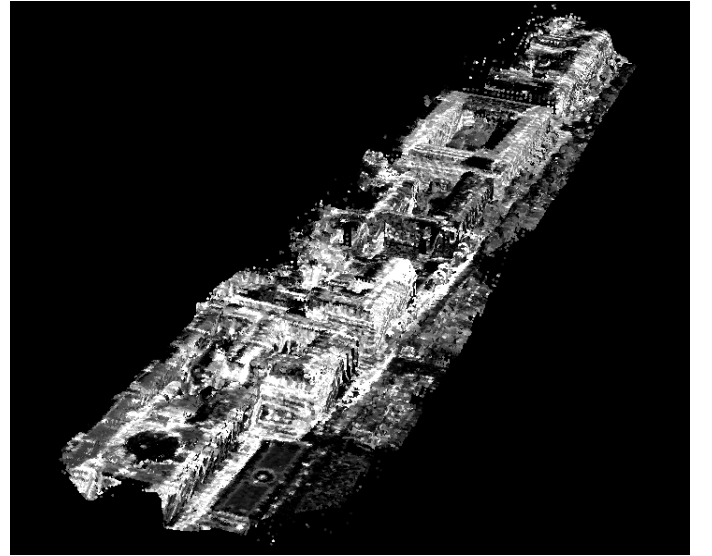


Figure 8. The building surface textured by the intensity of the scatterers from the TomoSAR point cloud. The brightness indicates the intensity of the scatterers in unit of dB.

IV. CONCLUSION AND OUTLOOK

This paper demonstrated a general framework of the fusion of spaceborne SAR data and street view images. Preliminary result using TerraSAR-X and Google map images were shown. By combining the distinguished ability of InSAR in measuring millimeter-level displacement and the rich semantic information of large amount of close-range street view images, a more detailed 3D building model with precise deformation

map that reveals unusual deformation behaviour on specific parts of the building can be produced.

The core of the demonstrated framework lies on the 3D reconstruction from both types of data, respectively. But there remains challenges. On one hand, even though city-scale 3D reconstruction can be obtained by D-TomoSAR technique, it is still a computationally expensive process, and a fairly large stack of well-coregistered SAR images is required. On the other hand, city-scale SfM point clouds from street view images are not easily obtainable, due to the limited availability and quality of online street view images. Our future work will be focused on fusing optical and TomoSAR point clouds of large urban areas, as well as using other optical image sources, such as social media image or consumer drone videos frames.

ACKNOWLEDGMENT

We gratefully acknowledge the support of the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No [ERC-2016-StG-714087], Acronym: *So2Sat*), and the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de).

REFERENCES

- [1] S. Agarwal *et al.*, "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, p. 105, Oct. 2011.
- [2] C. Wu, "Towards linear-time incremental structure from motion," in *3DTV-Conference, 2013 International Conference on*, 2013, pp. 127–134.
- [3] Y. Wang, X. X. Zhu, B. Zeisl, and M. Pollefeys, "Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 14–26, Jan. 2017.
- [4] X. Zhu and R. Bamler, "Very High Resolution Spaceborne SAR Tomography in Urban Environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, 2010.
- [5] X. Zhu and R. Bamler, "Tomographic SAR Inversion by L1-Norm Regularization -- The Compressive Sensing Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, 2010.
- [6] X. Zhu and R. Bamler, "Super-Resolution Power and Robustness of Compressive Sensing for Spectral Estimation With Application to Spaceborne Tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, 2012.
- [7] X. Zhu and R. Bamler, "Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, 2012.
- [8] Y. Wang, X. X. Zhu, R. Bamler, and S. Gernhardt, "Towards TerraSAR-X Street View: Creating City Point Cloud from Multi-Aspect Data Stacks," in *Urban Remote Sensing Event (JURSE), 2013 Joint*, 2013, pp. 198–201.
- [9] Y. Wang, X. Zhu, and R. Bamler, "An Efficient Tomographic Inversion Approach for Urban Mapping Using Meter Resolution SAR Image Stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1250–1254, 2014.
- [10] X. X. Zhu and R. Bamler, "Superresolving SAR Tomography for Multidimensional Imaging of Urban Areas: Compressive sensing-based TomoSAR inversion," *Signal Process. Mag. IEEE*, vol. 31, no. 4, pp. 51–58, Jul. 2014.
- [11] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, "Imaging Geodesy - Toward Centimeter-Level Ranging Accuracy With TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 661–671, Feb. 2011.
- [12] X. X. Zhu, S. Montazeri, C. Gisinger, R. F. Hanssen, and R. Bamler, "Geodetic SAR Tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 18–35, 2015.
- [13] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [14] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the World from Internet Photo Collections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 189–210, Nov. 2008.
- [15] M. Pollefeys *et al.*, "Detailed real-time urban 3d reconstruction from video," *Int. J. Comput. Vis.*, vol. 78, no. 2–3, pp. 143–167, 2008.
- [16] M. Pollefeys *et al.*, "Visual Modeling with a Hand-Held Camera," *Int J Comput Vis.*, vol. 59, no. 3, pp. 207–232, Sep. 2004.
- [17] J. Kang, M. Körner, Y. Wang, H. Taubenböck, and X. X. Zhu, "Building instance classification using street view images," *ISPRS J. Photogramm. Remote Sens.*
- [18] X. Cong, U. Balss, M. Eineder, and T. Fritz, "Imaging Geodesy — Centimeter-Level Ranging Accuracy With TerraSAR-X: An Update," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 948–952, Sep. 2012.
- [19] N. Adam, F. R. Gonzalez, A. Parizzi, and W. Liebhart, "Wide area persistent scatterer interferometry," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, 2011, pp. 1481–1484.
- [20] F. Rodriguez Gonzalez, N. Adam, A. Parizzi, and R. Brcic, "The Integrated Wide Area Processor (IWAP): A Processor For Wide Area Persistent Scatterer Interferometry," in *Proceedings of ESA Living Planet Symposium 2013*, Edinburgh, UK, 2013.
- [21] X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's Tomographic SAR Processing System," in *Urban Remote Sensing Event (JURSE), 2013 Joint*, 2013, pp. 159–162.
- [22] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, 2010.
- [23] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," p. 10.

